ISSN: 2321-7782 (Online)

Volume 3, Issue 1, January 2015

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Improving Network Performance by Differentiating DDoS Attacks from Flash Crowds

Kanchan H. Patil[1]
PG Student,
Department of Computer Engineering,
JSPM's RSCOE, Tathawade
Pune – India

Dr. Prof. A. B. Bagwan[2]
Department of Computer Engineering,
JSPM's RSCOE, Tathawade.
Pune – India

Dr. Prof. P. K Deshmukh[3]
Department of Computer Engineering,
JSPM's RSCOE, Tathawade.
Pune – India

Abstract: Todays internet system has various vulnerabilities and threats due to excessive use of it. These threats are increasing day by day among which one is DDoS attack. It is very critical task for today's defenders to detect DDoS attacks against flash crowds. Both DDoS attack and flash crowd cause surges of access to a server, but flash crowds are unexpected and legitimate. Main tool behind DDoS attack is botnet and botmasters try to disable detection strategy of DDoS attack by mimicking the patterns of flash crowd. So it is the most challenging task today to detect DDoS attack against flash crowd. It was found that the current attack flows are usually more similar to each other compared to the flows of flash crowds. Based on this, the flow correlation coefficient is used as a similarity metric to detect DDoS attack flows from genuine flash crowd flows.

Keywords: Threats, Vulnerabilities, DDoS attack, Botnet, Botmaster, Flash crowd, Flow similarities, Detection.

I. Introduction

Vulnerabilities to today's network system are increasing due to excessive and fast moving growth of internet. Attackers are continuously trying togain unauthorized access over it.One of the prominent threats to internet is Distributed Denial of Service (DDoS) attack. .In network system two events are there which flood the web server, namely DDoS attack and flash crowd. Distributed denial of service attacks contain malicious requests to subvert the normal operation of the website while flash crowds are due to a sudden, large surge in traffic to a particular Web site, created by legitimate requests. Botnets are main engines behind DDoS attacks, and they flood the victim webserver with service requests generated from many bots. Attack requests are similar in content to those generated by legitimate users.

Experienced botmasters attempt to disable detection strategy by mimicking the traffic patterns of flash crowds.So during the flash event, the main aim of server is to identify flash crowd attacks or DDoS attacks from genuine flash crowds. It was found that the similarities among the current DDoS attack flows are higher than that of a flash crowd flows. The reason for this phenomenon is that the number of live bots of a current botnet is far less than the number of concurrent legitimate users of a flash crowd.

This paper presents the proposed approach using the flow correlation coefficient as a metric to measure the similarity among suspicious flows to detect DDoS attacks from genuine flash crowds.

II. MOTIVATION

DDOS attacks have increased dramatically in recent years which are demonstrated by the survey [1] of the 70 largest internet operators in the world. Well experienced botmasters take advantage of various techniques to carry out their activities such as

- Code obfuscation, memory encryption [4] to disguise their traces
- Fresh code pushing for resurrection [5].
- Peer-to-peer implementation technology [6], [7], or flash crowd mimicking [8] in order to sustain their botnets.

Flash crowds are unexpected, but legitimate, dramatic surges of access to a server, such as breaking news. When web server is flooded with both DDoS attack and flash crowd requests, it is necessary for server to differentiate between these two types of requests.

Many previous research works have been carried out in attempt to differentiate DDoS attacks from flash crowds. These methods however cannot efficiently differentiate between DDoS attacks and flash crowds. Also most of these methods work properly at application layer only. Our proposed method uses flow correlation coefficient to differentiate between DDoS attacks and flash crowds and can effectively work on network layer also.

III. LITERATURE SURVEY

This section presents review of various research papers those are referred for study of discriminating DDoS attack from flash crowd. Also some facts about current botnets are covered in this section.

- Previous work [8], [9] focused on extracting DDoS attack features, and was followed by detecting and filtering DDoS attack packets by the known features. However, these methods cannot actively detect DDoS attacks.
- Discriminating DDoS Attack Traffic from Flash Crowd through Packet Arrival Patterns:

A behavior based detection that can discriminate DDoS attack traffic from traffic generated by real users is proposed. By using Pearson’s correlation coefficient, comparable detection methods [13] can extract the repeatable features of the packet arrivals.

- Discriminating DDoS Flows from Flash Crowds Using Information Distance:

This method employs abstract distance metrics, the Jeffrey distance, the Sibson distance, and the Hellinger distance to measure the similarity among flows to achieve goal. By comparing the three metrics [14] and found that the Sibson distance is the most suitable with accuracy around 65%.

- Distinguishing DDoS Attacks from Flash Crowds Using Probability Metrics:

This work propose a set of novel methods using probability metrics to distinguish DDoS attacks from Flash crowds and propose hybrid probability metrics[15] can greatly reduce both false positive and false negative rates in detection .

- Currently most popular defense against flash crowd attacks is the use of graphical puzzles to differentiate between humans and bots [10]. This method involves human responses and can be annoying to users.
- Xie and Yu tried to differentiate DDoS attacks from flash crowds at the application layer based on user browsing dynamics [11]. Oikonomou and Mirkovic tried to differentiate the two by modeling human behavior. These behavior-based discriminating methods work well at the application layer. However, it has not seen any detection method at the network layer, which can extend defense diameter far from the potential victim.
- There are a number of reports on the size and organization of botnets [7]. Bots are caught by honeypots and analyzed

thoroughly via inverse engineering techniques. Botnet infiltrations are further implemented to collect first-hand information about their activities [3], and even implemented a peer-to-peer-based botnet for research purposes.

- The attack tools are prebuilt programs, which are usually the same for one botnet. A botmaster issues a command to all bots in his botnet to start one attack session. This can be evidenced from the literature of botnet [4].
- The attack flows that are observed at the victim's end are an aggregation of many original attack flows, and such attack flows share a similar standard deviation as an original attack flow, and the flow standard deviation is usually smaller than that of genuine flash crowd flows. The reason for this phenomenon is that the number of live bots of a current botnet is far less than the number of concurrent legitimate users of a flash crowd.
- Rajab et al. recently reported that the live bots of a botnet is at the hundreds or a few thousands level for a given time point [16]. However, it is observed that the found on the Computer number of concurrent users of the flash crowds of World Cup 98 is at the hundreds of thousands level. Therefore, in order to launch a flash crowd attack, a botmaster has to force his live bots to generate many more attack packets, e.g., web page requests, than that of a legitimate user. As a result, the aggregated attack flow possesses a small standard deviation compared with that of a flash crowd.

IV. PROPOSED DIFFERENTIATION METHOD

Proposed differentiation method is based on flow analysis which uses feature of flow similarity to differentiate DDoS attacks from genuine flash crowds under current botnet size and organization, addressing the problem of differentiation at the network layer. Differentiation method computes correlation coefficient [2] which makes it delay proof and effective against explicit random delay insertion among attack flows. Differentiation algorithm works independently of specific DDoS flooding attack types.

V. SYSTEM BLOCK DIAGRAM AND DATA FLOW

A. *Block Diagram of System*

This differentiation system starts from captured packets as input to the system. Flow correlation coefficient is computed for the captured packets in the system. On the basis of this flow correlation coefficient value, the system will generate differentiation results and display differentiated attack (DDoS) packets from legitimate (flash crowd) packets.

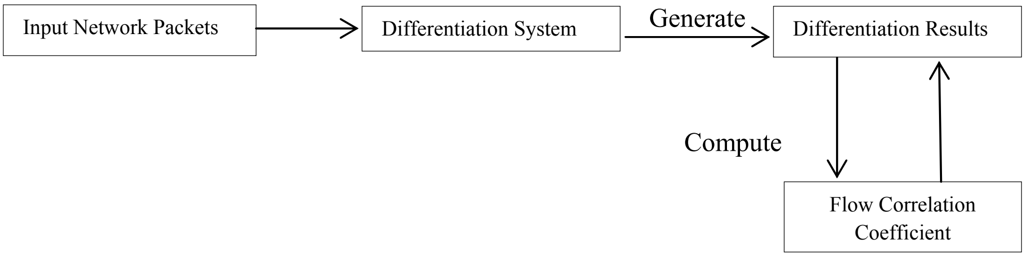


Fig1: Block Diagram of System

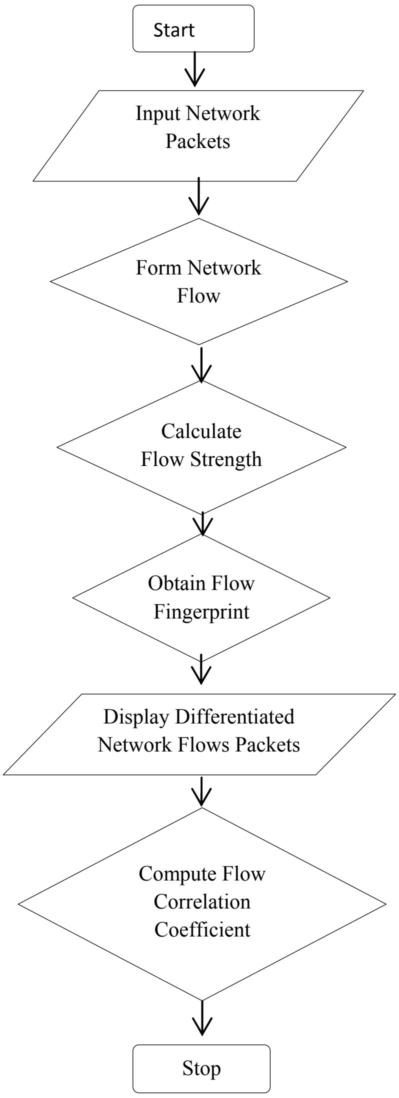
B. *Data Flow in System*

Figure2shows a Data flow in the differentiation system which is described by following steps.

1. Capture input network packets coming towards community network.
2. Form the network flows for each destination address.
3. Calculate flow strength of network flows.
4. Obtain flow fingerprint.
5. Compute flow correlation coefficient values between two flows with same length.

6. Display differentiated network flows packets.

7. Display result evaluation.



VI. Mathematical Model and Definitions

A. Mathematical Model

1. X [1……..M] = { X_1, X_{2,X_3} ....X_M }, M≥1, X is a network flow.

2. $X_i = \{x_{i_1}, x_{i_2}, x_{i_3} .... x_{i_N}\}$, $1 \le i \le M$ and $N \ge 1$, X_i is i^{th} network flow.

3. $X_{iK} = \{x_{i1}, x_{i2}, x_{i3} .... x_{iK}\}$, $1 \le K \le N$, Number of packets counted in k^{th} time interval.

4. X = <Protocol, Source IP, Source Port, Destination IP, Destination Port>

Protocol = <TCP, UDP>

Source IP =<32 bit Source IP Address>

Destination IP =<32 bit Destination IP Address>

Source Port =<16 bit Source Port number>

Destination Port = <16 bit Destination Port number>

5. $I_{X_i,X_j} = \{0, 1\}, 1 \leq i, j \leq M, i \neq j, X_i, X_j \in X,$ Similarity indicator.

$I_{X_i,X_j} = \{1\},$ DDoS attack Indicator.

B. Definitions

i. Network Flow:

For a given community network, cluster the network packets that share the same destination address as one network flow.Network flow is defined as,

$$X_i = \{x_i[1], x_i[2]........, x_i[N]\} \quad (1)$$

Where, X_i given network flow, N – length of given network flow, $x_i[k](1 \leq k \leq N)$ - represents the number of packets that we counted in the kth time interval for the network flow.

ii . Flow Strength:

Expectation of given flow is defined as flow strength of that flow.Flow strength represents the average packet rate of a network flow.

$$E[X_i] = \frac{1}{N}\sum_{n=1}^{N} x_i[n] \, (2)$$

Where, $E[X_i]$ - expectation of the flow (flow strength)

iii. Flow Fingerprint:

Flow fingerprint is the unified representation of the given network flow.

$$X_i^{'} = \{X_i^{'}[1], X_i^{'}[2]...., X_i^{'}[N]\} = \{\frac{x_i[1]}{N * E[X_i]}, \frac{x_i[2]}{N * E[X_i]}, ..., \frac{x_i[N]}{N * E[X_i]}\} \, (3)$$

Where, $X_i^{'}$ - fingerprint of flow X_i

On the basis of definition (2) and (3),a network flow and its fingerprint is related as,

$$X_i = N * E[X_i] * X_i^{'} \, (4)$$

Since $\sum_{k=1}^{N} X_i^{'}[k] = 1$

Correlation between the two flows is given as,

$$r_{X_i,X_j} = \frac{1}{N}\sum_{n=1}^{N} x_i[n] x_j[n] \quad (5)$$

It may be indicated zero correlation although the two flows are completely correlated with a phase difference. The definition therefore is modified as:

$$r_{X_i,X_j}[k]=\frac{1}{N}\sum_{n=1}^{N}x_i[n]x_j[n+k] \qquad (6)$$

Where, k (k = 0, 1, 2, ….., N-1) indicates the position shift of flow X_j.

4. Flow Correlation Coefficient

Flow Correlation Coefficient indicates similarity between two flows. Correlation coefficient of the two flows is defined as

$$\rho X_i, X_j[k]=\frac{rX_i, X_j[k]}{\frac{1}{N}\left[\sum_{n=1}^{N-1}x_i^2[n]\sum_{n=1}^{N-1}x_j^2[n]\right]^{1/2}}$$

For sampled M network flows $X_1, X_2, ..., X_M$.Obtain the flow correlation coefficient of any two network flows, $X_i(1 \le i \le M)$ and $X_j(1 \le j \le M, i \ne j)$. An indicator for the similarity is I_{X_i,X_j} of flow X_i and X_j, and which has only two possible values: 1 indicates DDoS attacks and 0 otherwise. Let δ be the threshold for the differentiation as,

$$I_{X\ ,X_j}=\begin{cases}1_, & \rho X_i, X_j[k] \ge \delta, \\ 0_, & \text{otherwise,}\end{cases} \qquad (7)$$

Where, $1 \le i, j \le M$ and $i \ne j$

In a community network, there may have two suspected flows. Therefore pair wise comparisons can be conducted to derive the result.

VII. Definition Algorithm

Our differentiation system requires captured network packets as an input and using differentiation algorithm it differentiates DDoS attack flows from legitimate flash crowd flows.

- Algorithm for Network Flow Differentiation

```
1.  Start
2.  Initialize n, δ                          // n- is packet sample size, δ -discrimination threshold.
3.  Identify  X , m                          // X- network flow, m– number of destination addresses
4.  Until  sample size>=n do
    $X_i = \{x_i[1], x_i[2]…..,x_i[n]\}$     // i(I>= m), $X_i$ – $i^{th}$ network flow, x-network flow packet
5.  Calculate
    $FS[X_i]=\frac{1}{n}\sum_{n=1}^{n}x_i[n]$     // FS -    flow strength of Xi
6.  Calculate
    $FF[X_i]=\{\frac{x_i[1]}{n*E[X_i]},\frac{x_i[2]}{n*E[X_i]},.....,\frac{x_i[n]}{n*E[X_i]}\}$     //FF – flow fingerprint of Xi
7.  Go to step 3.
```

```
8.  Until i>= m do
9.  Calculate
    $r_{X_i,X_j} = \frac{1}{n}\sum_{n=1}^{N} X_i[n]X_j[n]$
    //r –Correlation between the two flows
    $FCC[X_i, X_j] = \frac{rX_i, X_j}{\frac{1}{n}\left[\sum_{n=1}^{n-1} x_i^2[n]\sum_{n=1}^{n-1} x_j^2[n]\right]^{1/2}}$
    //FCC – Flow correlation coefficient between Xi and Xj
10. Compare between two flows
        If (FCC [$X_i$,$X_j$] >= δ )
        {
                DDoS attack flows
        }
        Else
        {
                Flash crowd flows
        }
11. Display differentiated packets of network flows
12. Stop
```

VIII. Results

Resultant graph for input network packet files is as shown in figure1 below. Our proposed method calculates pair wise flow correlation coefficient value for different network flows in input network packet files. The flow correlation coefficient value is compared with threshold value which is defined in between 0 and 1(considered as 0.5 for comparison). If the correlation coefficient is greater than threshold then packets are considered as attack (DDoS) packets otherwise legitimate (flash crowd) packets.

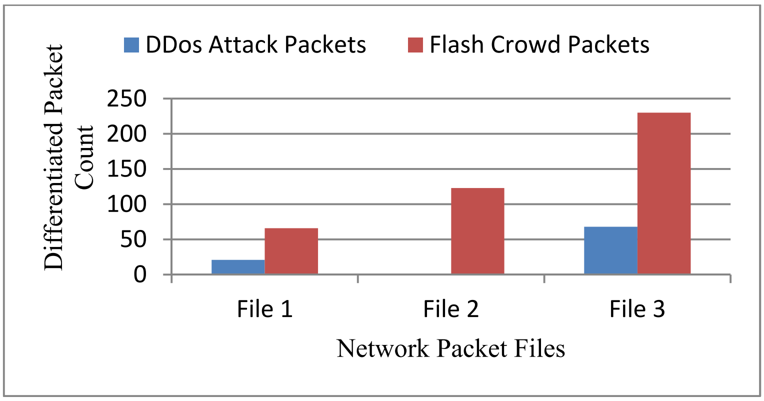


Figure3: Resultant Graph Showing Differentiated Attack (DDoS) Packets from Legitimate (flash crowd) Packets.

IX. CONCLUSION

Proposed differentiation method tried to differentiate distributed denial of service attacks from genuine flash crowds which is most challenging problem today .It found that under the current conditions of botnet size and organization, DDoS attack flows have more similarity than genuine flash crowd flows. So our method used flow correlation coefficient as a metric to measure similarity among network flows. Result confirmed differentiation between DDoS attack flows and genuine flash crowd flows.

Future work will focus on possibility of organizing a super botnet, with a sufficiently large number of live bots which can beat the proposed method. Secondly, if the attacker is known with the proposed strategy then it is necessary to explore actions which there should have to take against attacker's actions.

References

1. Arbor, "IP Flow-Based Technology," http://www.arbornetworks.com, 2011.
2. Shui Yu, WeijiaJia, Song Guo, Yong Xiang, and Feilong Tang "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 23, NO. 6, JUNE 2012.
3. B. Stone-Gross, M. Cova, L. Cavallaro, B. Gilbert, M. Szydlowski,R. Kemmerer, C. Kruegel, and G. Vigna, "Your Botnet Is My Botnet: Analysis of a Botnet Takeover,"Proc. ACM Conf. Computer Comm. Security,2009.
4. N. Ianelli and A. Hackworth, "Botnets as Vehicle for Online Crime,"Proc. 18th Ann. First Conf.,2006.
5. C.Y. Cho, J. Caballero, C. Grier, V. Paxson, and D. Song, "Insights from the Inside: A View of Botnet Management from Infiltration," Proc. Third USENIX Conf. Large-Scale Exploits and Emergent Threats:Botnets, Spyware, Worms, and More (USENIX LEET),2010.
6. T. Holz, M. Steiner, F. Dahl, E. Biersack, and F.C. Freiling,"Measurements and Mitigation of Peer-to-Peer-Based Botnets: A Case Study on Storm Worm,"Proc. First Usenix Workshop Large-Scale Exploits and Emergent Threats (LEET),2008.
7. M. Bailey, E. Cooke, F. Jahanian, Y. Xu, and M. Karir, "A Survey of Botnet Technology and Defenses,"Proc. Cybersecurity Applications and Technology Conf. for Homeland Security,2009.
8. J. Jung, B. Krishnamurthy, and M. Rabinovich, "Flash Crowds and Denial of Service Attacks: Characterization and Implications for CDNs and Web Sites,"Proc. 11th Int'l Conf. World Wide Web (WWW),pp. 252-262, 2002.
9. G. Carl, G. Kesidis, R. Brooks, and S. Rai, "Denial-of-Service Attack-Detection Techniques,"IEEE Internet Computing, vol. 10,no. 1, pp. 82-89, Jan./Feb. 2006.
10. S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-Sale:Surviving Organized DDoS Attacks that Mimic Flash Crowds (Awarded Best Student Paper)," Proc. Second Symp. Networked Systems Design and Implementation (NSDI '05),2005.
11. Y. Xie and S.-Z. Yu, "A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors,"IEEE/ACM Trans. Networking,vol. 17, no. 1, pp. 54-65, Feb. 2009.
12. G. Oikonomou and J. Mirkovic, "Modeling Human Behavior for Defense against Flash-Crowd Attacks,"Proc. IEEE Int'l Conf.Comm.,2009.
13. TheerasakThapngam, Shui Yu, Wanlei Zhou and GlebBeliakov "Discriminating DDoS Attack Traffic from Flash Crowd through Packet Arrival Patterns" The First International Workshop on Security in Computers, Networking and Communications pno 969-974.
14. Shui Yu, TheerasakThapngam, Jianwen Liu, Su Wei and Wanlei Zhou "Discriminating DDoS Flows from Flash Crowds Using Information Distance" International Conference on Network and System Security pno 351-356.2009.
15. Ke Li, Wanlei Zhou, Ping Li, Jing Hai and Jianwen Liu "Distinguishing DDoS Attacks from Flash Crowds Using Probability Metrics" Third International Conference on Network and System Security pno: 9-17 .2009.
16. M.A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My Botnet is Bigger than Yours (Maybe, Better than Yours): Why Size Estimates Remain Challenging,"Proc. First Conf. First Workshop Hot Topics in Understanding Botnets (HotBots '07),2007.